Handwriting for Efficient Text Entry in Industrial VR Applications: Influence of Board Orientation and Sensory Feedback on Performance

Nicolas Fourrier, Guillaume Moreau, Mustapha Benaouicha, Jean-Marie Normand

Handwriting for efficient text entry in industrial VR applications: influence of board orientation and sensory feedback on performance.

Nicolas Fourrier, Guillaume Moreau, Mustapha Benaouicha, and Jean-Marie Normand

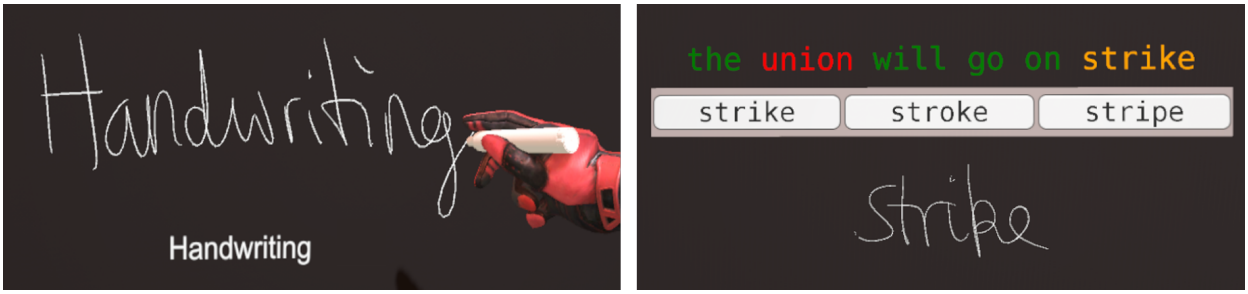

Fig. 1: The user can write words by hand on a board in the virtual environment. Handwriting recognition technology converts the handwritten word into usable text data (left). During the user study, participants are presented with phrases they have to write one word at a time (right). The yellow word is the selected word. Green words are correctly recognized and red words are not. Users are presented with multiple recognition results they can choose from.

Abstract—Text entry in Virtual Reality (VR) is becoming an increasingly important task as the availability of hardware increases and the range of VR applications widens. This is especially true for VR industrial applications where users need to input data frequently. Large-scale industrial adoption of VR is still hampered by the productivity gap between entering data via a physical keyboard and VR data entry methods. Data entry needs to be efficient, easy-to-use and to learn and not frustrating. In this paper, we present a new data entry method based on handwriting recognition (HWR). Users can input text by simply writing on a virtual surface. We conduct a user study to determine the best writing conditions when it comes to surface orientation and sensory feedback. This feedback consists of visual, haptic, and auditory cues. We find that using a slanted board with sensory feedback is best to maximize writing speeds and minimize physical demand. We also evaluate the performance of our method in terms of text entry speed, error rate, usability and workload. The results show that handwriting in VR has high entry speed, usability with little training compared to other controller-based virtual text entry techniques. The system could be further improved by reducing high error rates through the use of more efficient handwriting recognition tools. In fact, the total error rate is 9.28% in the best condition. After 40 phrases of training, participants reach an average of 14.5 WPM, while a group with high VR familiarity reach 16.16 WPM after the same training. The highest observed textual data entry speed is 21.11 WPM.

Index Terms—Virtual reality, handwriting, text entry, industry.

1 Introduction

Text entry in Virtual Reality (VR) is an increasingly common task used in social, collaborative, and industrial applications. Whether for authentication, communication, or more specific operations such as filtering databases, VR users face short yet frequent typing tasks. In industrial environments, there are many barriers to VR adoption such as integration difficulties in current workflows or lack of expertise which can be offset by productivity and quality gains [2, 31]. However, current VR text input techniques are less efficient than their traditional computer counterparts [10], even for techniques that rely on the use of external physical keyboards [16]. This loss in efficiency in VR limits productivity gains and makes data entry, which is a recurring and performance critical task in industry, an additional barrier to adoption. There is currently a need for efficient, easy-to-learn text entry in VR that does not generate any frustration. Text entry systems that rely on external devices, such as actual keyboards, may not only be inconvenient and costly due to the technologies they use like hand and finger-tracking, they can also break immersion and reduce usability of VR [10]. Virtual keyboards solutions are far more flexible but their performance, typing speed and learnability do no meet industrial requirements. Some techniques are based on entirely different principles such as gestures [53] or voice recognition [34]. While they can sometimes rival virtual keyboards in terms of entry speed, they are not suitable for all use-cases, including industrial applications, for many reasons including privacy or learnability. As such, there is still a need to find an innovative way to input textual data that is efficient and easy-to-learn while not generating frustration. Indeed, none of the current techniques fulfill all three criteria simultaneously. For industrial purposes, it is not only the performance of a system that needs to be studied, but also its usability [22], which is not a primary concern in most VR data entry studies.

In this paper, we present a handwriting recognition system that allows VR users to manually write text on a virtual board. This text is then turned into usable textual data. The system is designed to maximize efficiency, learnability and to make the writing task as natural as possible through the study of different types of writing boards and the use of sensory feedback. A within-participant user study is conducted

- *Nicolas Fourrier is with Segula Technologies, Naval and Energy Engineering Research and Innovation Unit and École Centrale Nantes, AAU-CRENAU, UMR 1563 and LS2N-PACCE, UMR 6004, France. E-mail: nicolas.fourrier@segula.fr.*
- *Guillaume Moreau is with IMT Atlantique - Lab-STICC UMR 6285, France. E-mail: guillaume.moreau@imt-atlantique.fr.*
- *Mustapha Benaouicha is with Segula Technologies, Naval and Energy Engineering Research and Innovation Unit and Cherbourg University Laboratory of Applied Sciences LUSAC, University of Caen Normandy, France. E-mail: mustapha.benaouicha@segula.fr.*
- *Jean-Marie Normand is with Nantes Université, École Centrale Nantes, AAU-CRENAU, UMR 1563 and LS2N-PACCE, UMR 6004, F-44000 Nantes, France. E-mail: jean-marie.normand@ec-nantes.fr.*

with 40 participants to determine the best implementation of the system and its performance. The best performing implementation (slanted writing board with sensory feedback) achieved an average speed of 14.3 words per minute (WPM) and 9.28% total error rate (TER). In this implementation, the best participant reached a writing speed of 21.11 WPM. After 40 training phrases, the average writing speed is 14.5 WPM and 16.16 WPM for VR experienced participants. Our proposed method only relies on the use of a single controller and could be further optimized using state-of-the-art handwriting recognition, as reducing error rate would contribute greatly to improve writing speed. Nevertheless, we still need a long-term learnability evaluation and a direct comparison with state-of-the-art entry systems of our system.

In summary, our contributions in this work are (1) a user study to determine the performance and usability of a text entry handwriting system in VR depending on board orientation and sensory feedback, (2) a set of metrics adapted to word-by-word systems based on character-level metrics definitions and (3) a list of criteria VR text entry techniques must meet in order to be usable in industrial settings.

The remainder of this paper is as follows: Sect. 2 presents a review of state-of-the-art virtual text entry methods in VR, Sect. 3 details the design rationale and Sect. 4 describes our handwriting recognition system for text entry. Our user study that aims to determine the best writing conditions and overall performance of the system is presented in Sect. 5. Finally, in Sect. 6 we discuss our findings as well as future work leads before concluding in Sect. 7.

2 Related work

Text entry techniques in VR can be divided into two groups: physical or virtual techniques. Physical techniques rely on the use of real keyboards or external tools. Virtual techniques, on the other hand, only rely on a traditional VR setup such as a tracked Head-Mounted Display (HMD) and controllers [10]. Overall, physical techniques show considerably higher typing speed than virtual ones. However, they are often not compatible with VR applications that require moving around. Moreover, keyboard users often rely on looking at the keyboard or at their hands, which is difficult when immersed in VR [24]. Virtual text entry techniques, while often displaying considerably slower typing speed, can be used in any VR scenario without needing additional tools, sometimes without even using controllers. These virtual techniques often also rely on displaying fairly large user interface element that may a problem in spatially constrained virtual environments (VEs).

This section will focus solely on virtual techniques as physical ones are not suited for the industrial use-case we study. Indeed, we focus on standard HMD and controllers VR setups that do not rely on external devices. Virtual techniques can be divided in two categories: selection-based techniques that use keyboard-like input systems, and innovative techniques that rely on entirely different principles such as gestures, voice recognition or handwriting.

2.1 Virtual keyboard-based text entry

Head-based Head-based techniques have been developed for typing text in VR. Head movement can be used to point at a virtual keyboard using different interaction strategies. Lu et al. [28] compare dwell, eye blink, and neck forward interactions. Blinking is shown to be the most efficient interaction type and reaches 13.47 WPM. Obviously, this requires eye-tracking, a feature that producers of VR devices are increasingly embracing. Head and eye movements should be used with caution in VR as they may cause sickness [7]. Yu et al. [51] also explore head-based typing techniques and show that using a gesture-based swipe keyboard results in higher performance. Users achieve 24.73 WPM after typing 80 phrases with an "acceptable" subjective rating score of fatigue (2.61 on a 1-5 scale). The downside of such systems is that they cannot be used to type words that are not in the dictionary such as passwords or technical data. Gaze-based text entry, using eye movements to aim at a keyboard, has also been studied. In [36], Rajanna et al. shows that such a system achieved 10.15 WPM with a 7% error rate, provided that users should not experience motion in their field of view as it can induce strain and motion sickness.

Controller-based Head pointing techniques have also been directly compared to raycast (or point-and-click) and drum-like methods by Boletsis et al. [4] and Speicher et al. [40]. Drum techniques are highly efficient as they achieve 21 WPM, followed by raycast with a performance of $\sim$16 WPM and head pointing reaching $\sim$10 WPM. Drum-based and pointing techniques can be found physically demanding as some users commented that using two controllers can be tiring. To the best of our knowledge, single controller versions of these virtual keyboard techniques have not been extensively studied even if dual raycast "clearly outperforms" single raycast according to Outerelo [33]. Controller-based pointing systems can be combined with a word-gesture text entry to improve performance as is shown by Chen et al. [8] where users achieve 16.43 WPM after only writing 10 phrases.

Mid-air typing Using hand-tracking for typing on virtual keyboards is a promising concept that performs best when aligning the virtual keyboard with a physical surface to provide passive haptic feedback [11]. High typing speeds of 55.6 WPM are reported in this condition. Mid-air keyboard typing can be improved by personalizing keyboards sizes and layouts [37]. Hand gestures can be used to switch efficiently between keyboard types to improve special character entry [38]. Historically, mid-air typing relies on expensive extra devices such as gloves [5] or sensors [50]. However, hand-tracking is an increasingly common built-in feature in Mixed Reality devices so mid-air typing might get more popular in applications that do not need controllers.

Non-traditional layouts Selection-based techniques not based on typical QWERTY-layout keyboards also exist. Indeed, many studies investigate circular and radial layouts. RingText [48] users can choose characters from a circular layout keyboard with head movements. Using some amount of auto-completion, this system achieves 11.30 WPM after 60 minutes of training for novice users. PizzaText [52] uses two thumbsticks to manually choose characters on a keyboard shaped as a slice of pizza. It reaches 8.59 WPM for novice users using a regular gamepad, not tracked VR controllers. HiPad [20] uses a keyboard layout similar to PizzaText with different interactions such as moving the controller in the virtual environment or using a touchpad. Novice users with 50 training phrases achieve 13.57 WPM in the VE condition. Flower Text Entry uses 3D hand translation to select letters in a flower-shaped keyboard [27]. Novice users reach 17.65 WPM after 80 phrases of training. This technique requires some amount of physical space, which might not be suitable for crowded environments. Yanagihara et al. study a cubic keyboard in which users can select keys by moving the controller in a $3 \times 3 \times 3$ spatial grid [49]. Slowing down movement below a certain threshold validates the currently selected key. The average entry speed of this technique is 21.59 WPM in a pilot study. BlueTap maps letters onto fingers that can be selected by taping different parts of the fingers with the thumb [9]. It reaches 10 WPM in pilot studies but used external sensors such as a wrist-worn camera which might limit freedom of movement in VR.

2.2 Virtual non keyboard-based techniques

Voice & gesture Some data entry techniques do not rely on selection-based paradigms. Voice recognition is an alternative that can reach high entry speeds but cannot be used in all situations. It may be difficult to use such systems in collaborative settings or when privacy is needed [16]. Moreover, speech-based error correction may prove difficult [45]. Users of SWIFTER, a voice recognition input system, reach an average input speed of 23.6 WPM [34]. Alternatively, many investigate gestures as a data entry method. 2-Thumbs Typing is a gesture-based method that uses touchpads to draw lines [53]. It requires two touchpads and learning gestures for each character. Users achieve 8.5 WPM with extensive training but the systems presents the advantage of being comfortable and not relying on sight. Hand gesture is also explored to draw letters in mid-air with fingers [32]. Typing speed is not measured in this study but the algorithm achieves 96.12% recognition rate. A similar system using disturbances in WiFi signal for tracking hand movement achieves 12.5 letters per minute (about 2.5 WPM) and 74.84% recognition rate [14].

Handwriting techniques are studied as early as 1998 by Poupyrev et al. [35]. Their Virtual Notepad uses a tracked tablet and a pen to let

users draw on it. The system is not designed for text entry but rather for taking notes or annotating scenes. In 2009, Macias et al. [15] compare a character-by-character handwriting entry tool to other systems but it scores low on typing speed and high on error rate. In 2017, in a system that lets users draw characters on boards from a long distance by pointing, it is noted that handwriting is a fun and engaging way to input textual data, but that it performs poorly in terms of typing speed and error rate [12]. More recently, Venkatakrishnan et al. [44] study the best conditions for handwriting in VR, using state-of-the-art recognition technology that performs way better than previous attempts. They find that writing on a flat surface by directly putting a virtual pen in contact with the writing surface performs best compared to curved surfaces, controller-pointing writing and finger-tracking writing. Users achieve a typing speed of around 10 WPM with near perfect accuracy. Hsu et al. [19] highlight the need to further study and develop tools for fine motor tasks in VR such as writing.

2.3 Text entry task and metrics

Task Text entry systems are evaluated by doing a transcription task where participants must type given phrases as fast and as accurately as possible. Different phrase sets exist for evaluating text entry. Some focus more on internal validity i.e., they aim that variation in dependent variables does not originate from the phrases themselves, while some focus more on external validity i.e., they aim for the phrases to be representative of real-life applications of text entry systems. Ideally, a phrase set should respect those two criteria [25]. The most commonly used phrase set since its release in 2003 is the MacKenzie phrase set [29]. It is composed of 500 phrases that contain no punctuation and a low amount of uppercase characters. It has high correlation with the English language when it comes to character frequency. In VR text entry studies, the MacKenzie phrase set seems to be exclusively used [4, 8, 20, 24, 27, 28, 36, 40, 51, 52]. During the text entry task, users can be instructed not to correct errors, invited to correct errors or forced to correct errors. These instructions can have a major impact on entry speed and error rate metrics and it is important to choose wisely [1].

Text entry speed Measuring speed in text entry tasks can be done using the WPM metric [30] defined as:

$$WPM = ((|T|-1)/S) \times 60 \times 1/5\,, \tag{1}$$

where $|T|$ is the length of the transcribed string, and S the time it took to enter said string from first to last entry [30]. One character is subtracted from $|T|$ since the timer often starts after validating the first character. This is not always taken into account [52]. WPM is the most commonly used metric for text entry speed [1]. Adjusted WPM (AdjWPM) is similar but compensates for errors in the final transcribed text [30]. Other metrics exist such as KeyStrokes per Second (KSPS) and Gesture per Second (GPS) that aim to provide indication to what happens during text entry, while WPM only looks at the final transcribed text. AdjWPM, KSPS, and GPS are not commonly used [1].

Error rates Error rate is usually reported by using two metrics: Total Error Rate (TER) that includes both corrected and uncorrected errors, and Not Corrected Error Rate (NCER) that only includes uncorrected errors that are present in the final transcribed string. NCER is often only called Error Rate (ER). Reporting both these metrics is often necessary because ER alone does not tell any information about the effort made to correct errors during the transcription task. The definitions of these metrics, in percentage (%), are as follows:

$$ER = (INF/|T|) \times 100\,, \tag{2}$$

$$TER = ((IF+INF)/(C+INF+IF)) \times 100\,, \tag{3}$$

where INF is the number of Incorrect Not Fixed characters, IF is the number of Incorrect Fixed characters, and C is the number of Correct characters, where fixed means corrected. The Levenshtein Minimum String Distance algorithm (MSD) is often used to compute ER as:

$$MSD\,ER = (MSD(P,T)/MAX(|P|,|T|)) \times 100\,, \tag{4}$$

where P is the presented text and T the transcribed text [39]. This algorithm aims to find the number of required transformation to change the transcribed text into the presented text in terms of deletions, insertions or subtractions. ER and MSD ER are "almost equivalent" according to Arif et al. [1] and can be used interchangeably. Dube et al. state that most studies report ER and TER but that they may compute ER differently [10]. Indeed, there seems to be some ambiguity when it comes to the definition of ER and the computation of INF. Another metric rarely used in VR text entry studies is the KeyStroke per Character (KSPC). This metric, categorized as an efficiency measure by MacKenzie et al. [30], aims to determine the mean effort made to enter each character, taking into account errors. Arif&Stuerzlinger show that there is no significant difference between a forced and recommended correction condition when it comes to KSPC and TER as users usually correct errors as soon as they spot them if given the chance [1].

This section shows the difficulty of effectively comparing systems as the metrics used are not always the same. Even when they are, they still may not be computed in the same way. In the case of handwriting, which is a word-by-word entry process, there is also a need to define metrics that remain comparable to keyboard-like character-by-character metrics. A word uncorrected error rate is mentioned in SWIFTER [34], a voice recognition input system, but this metric alone is not sufficient to highlight the effort made to correct words during the text entry task and no clear definition is given. There is a need to expand these word-level definitions and to properly state how to compute them.

3 Design Rationale

As we have seen in Sect. 2, no current VR text entry system appears to be well suited for repeated typing activities that need to be handled quickly and without frustration. This is especially true when considering our industrial use case of VR i.e., using a typical HMD setup with controllers in an empty room. Indeed, physical text-entry techniques impose spatial and technological constraints. The best performing virtual keyboard-based techniques reach about 25 WPM [4, 51] which is far lower than average typing speeds of experienced typists of 67 WPM on physical keyboards [24]. Innovative virtual techniques such as using hand-tracking for mid-air typing can reach 55 WPM in VR [11] but rely on complex and expensive technologies to prevent users from holding a controller in their hand while typing. Techniques that are not selection-based can rival in some circumstances with virtual keyboard-based techniques like voice recognition [34]. However, this technology is not well suited for industrial use for many reasons: it is noise-sensitive, does not work well in collaborative settings, and there may be privacy concerns. Handwriting is an alternative natural method of text entry that has potential as there does not seem to be any possible performance breakthroughs with virtual keyboards. New innovative paradigms need to be tested for VR. The first studies of handwriting in VR as a data entry tool describe some conditions under which it can perform adequately [44]. However, these studies have not focused on reaching high text entry performance.

Aiming to design a text entry system in VR that is efficient, easy-to-learn, that does not generate frustration and is usable in industrial environments leads us to develop a handwriting system that respects the following design rationales for these aspects.

3.1 Efficiency

Most current industrial computer-aided design tasks are performed on computers but many have studied the use of VR to replace or assist users [17,23,26,41,43,46]. Indeed, VR presents many advantages when it comes to 3D data visualization and collaboration. However, design and industrial tasks are mingled with data entry situations: filtering databases, naming parts, settings dimensions, etc. This makes it difficult to use VR for industrial tasks because overall productivity and efficiency are limited by basic interactions in VR such as inputting textual data.

Using a single controller, we want to design an efficient system that does not rely on traditional selection-based techniques for data entry because their performance is not up to par with industrial standards. Handwriting comes as the second most natural way to efficiently create textual data when one is not able to use a computer keyboard. Indeed,

it is estimated the average writing speed of young adults is around 40 WPM [42] in note-taking or fast writing situations. Reaching such speeds in VR would be a great improvement and it mostly depends on two factors: (i) how well the writing task is translated in VR and (ii) how well the recognition software performs. We decide to use an ink recognition software that, contrary to optical character recognition (OCR) algorithms, also uses a temporal information to recognize handwriting. This can help differentiating words thanks to the way letters are drawn without only relying on the final handwriting picture. Additionally, ink recognition software is mainly designed for handwriting recognition while OCR usually performs better on printed or typed text. Recognition performance is also improved by letting users choose from some of the most likely recognized words, which can change the output without having to rewrite the whole word. Efficiency is maximized by letting users write complete words and validate them with a simple button press. Button presses are indeed more efficient than dwelling or time-based validation methods [51]. A single controller button press is necessary per word. It could be argued that writing whole sentences would be more efficient but the writing board would have to be scaled accordingly. This would take up too much space in the virtual environment as handwriting tends to be larger in VR than in real life based on preliminary trials. Moreover, writing whole sentences will likely negatively impact correcting parts of a sentence. We believe writing at a word-scale is a good compromise between ease-of-use, flexibility and having more control over the output of the recognition software.

3.2 Ease-of-learning & familiarity

Making the VR writing task as natural as possible is crucial for performance and efficiency because familiarity will help learning the system. The time it takes to learn and master a system is key when it comes to its adoption in industrial settings. Performance is expected from the get-go and this must be accounted for. Handwriting is a task virtually everyone is trained for from a young age. As such, if VR handwriting is similar to real handwriting, the system should be easy-to-learn. We decide to focus for this study on standing-up handwriting with no physical supports as that would cover more potential industrial use cases of the system. A familiar task that fits with this use case is writing on whiteboards or blackboards. We focus on making the virtual handwriting task as natural as possible to maximize efficiency and learnability. The main difference between virtual and real writing is the lack of physical support, as one cannot rest on a virtual board. To enhance familiarity, we decide to present users with a blackboard type writing surface which can be written on with a virtual pen by direct contact [44]. A particular focus on sensory feedback (visual, audio, haptic) is needed for the task to feel natural. Indeed in a real handwriting task, it is expected that the pen will write on a surface with varying width depending on the force applied, without the pen and the board intersecting, and that the pen and hand cast shadows on the writing surface (see Fig. 2). It is also expected that the pen makes some sound when writing. Finally, haptics are also involved when writing most notably through the contact with the writing surface and vibrations. We hypothesize all these sensory feedback can be simulated and integrated into the handwriting tool in VR to improve familiarity, learnability and by extension, performance.

4 Implementation Details

With the previous design rationale and industrial use-case in mind, we design a VR input solution as a single-controller textual data entry system based on handwriting recognition technology.

4.1 Handwriting

According to Venkatakrishnan et al. [44] findings regarding handwriting in VR, we develop a system that uses a planar surface to write on by direct contact of a pen (see Fig. 3). We use a blackboard with white ink to reduce eye strain compared to using a whiteboard with black ink [13]. A particular attention is given to developing sensory feedback as stated in Sect. 3 as we hypothesize it will make the task more natural and facilitate learning. Visual feedback (see Fig. 2) is composed of (i) the varying width of the writing line depending on the force applied, (ii) the collision between the pen and the board, and (iii) a circle that changes size and color depending on the force applied. The writing force is computed as a function of the depth of the intersection between the pen and the board. When collision is on, we introduce a small visuo-proprioceptive conflict by disaligning the real and the virtual hands. The collision is not physics-based, it is only a displacement in the contact normal direction. The presence of a non-animated virtual hand is an additional visuo-motor feedback, that also acts as a visual feedback since the hand projects an additional shadow on the board. Active haptic feedback is provided by controller vibrations which vary depending on the writing force. Finally, audio feedback is given through different sounds: when first touching the board with the pen and when writing. The sounds are recorded from real writing on a board.

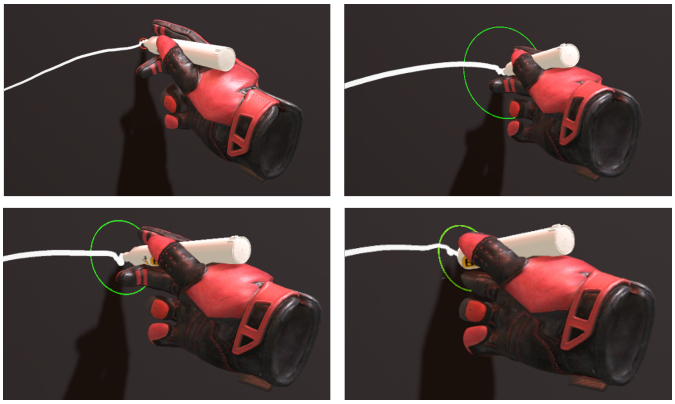

Fig. 2: Visual feedback of the handwriting system. Line width, circle size and color vary depending on the applied force (top). Collision between pen and board can be activated or not (bottom). When it is not, the pen and hand clip through the board (bottom-right).

4.2 Recognition

An offline handwriting recognition software called WritePad SDK [47], claiming to reach 96% accuracy, is adapted for integration in Unity. Its customizability, flexiblity, response times and the privacy provided by offline tools were chosen over using online handwriting recognition APIs. WritePad SDK expects not only a list of points, but also a temporal information. More precisely, this recognition software is given a temporally ordered list of 2D coordinates (x, y). Transforming the 3D positions of the contacts between the pen and the writing surface into the 2D ink data expected by the recognizer is trivial since our writing surfaces are planar. Using handwriting recognition lets us turn ink strokes into an exploitable textual data as described in Fig. 1. The goal of our study is not to evaluate the recognition software, but to provide insight on the usability of handwriting as a text entry tool in VR. As such, the system is designed so that the recognizer can easily be changed. The recognizer returns a list of possibly recognized words and an associated confidence score. When users write a word and validate it, they are also given the opportunity to choose between the three best recognition alternatives as seen in Fig. 1. Confidence score is usually too low for the remaining alternatives which is why we settle on displaying only three. There is no autocompletion and alternatives only appear after a word is validated. The recognizer can use dictionaries to help recognition but it remains possible to write technical data, like product references, by either using custom dictionaries or disabling dictionaries altogether. It is possible to choose between full sentence or individual word recognition. We choose word-by-word writing as it makes it easier to correct errors. Number and special character recognition can also be turned on which is necessary for industry.

4.3 Interaction

Steam Index controllers are used as their handling lets us simulate holding a pen. Other controllers are compatible. The user may validate words to start the recognition process by pressing a button under their thumb. This interaction is analog to pressing a key on a keyboard so

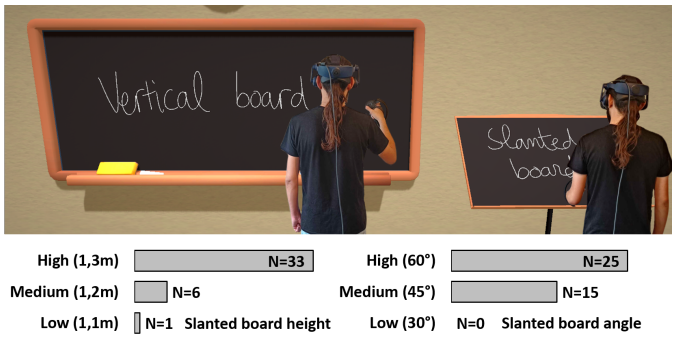

Fig. 3: Two boards are studied: a vertical one and a slanted one. The slanted board's height (1.1, 1.2, 1.3 in Unity spatial units) and angle (30°, 45°, 60°) can be configured by users. A high angle means a more vertical slanted board. Histograms show a breakdown of user choices for slanted board height and angle.

we add a space after each validated word in the transcribed text. After validating a word, users are presented with recognition alternatives that show up as virtual interactive buttons on the writing board (see Fig. 1). The user can replace the recognized word by one of the alternatives by pressing the associated virtual button with the pen. This interaction does not involve physically pressing a button on the controller. Users can erase the board by pressing a button on the controller which removes all ink present on the writing board. The user can select the previous word to rewrite it by moving the joystick to the left.

5 User Study

We determine the optimal system design for handwriting in VR, in terms of board type, sensory feedback but also performance and usability through a user study. In this study, the recognizer uses the built-in English dictionary and special characters and numbers are disabled since the MacKenzie phrase set does not contain them. The system is set to recognize one word at a time. Considering our industrial use case i.e., users wearing an HMD in an empty room, we restrict this study to a standing condition (see Fig. 3). We consider two board types. A vertical board, similar to a blackboard writing situation, is hypothesized to ease learning thanks to familiarity. A slanted and lower board, is believed to reduce physical load at the cost of familiarity. Indeed, writing on a virtual vertical board can be tiring due to the lack of physical support. A slanted and lower board does not require users to raise their arm for writing. We do not consider horizontal surfaces because they did not work well in early tests of our standing-up writing use case. We also study the impact of sensory feedback on writing in VR. Two conditions are studied: (i) the presence or (ii) the absence of all (visual, haptic, auditory and hand visibility, see Sect. 4) sensory feedback, which are hypothesized to have effects on error rates and writing speeds.

5.1 Hypotheses

We have formulated five hypotheses for this user study:
H.1. Writing on a vertical board will be faster than writing on the slanted board since users are more familiar with a vertical board.
H.2. Writing with sensory feedback will be faster than writing without since it will make the task more natural.
H.3. Writing on a slanted board will result in higher error rates since it will be harder to estimate the distance between the pen and the board than with the vertical board.
H.4. Writing without sensory feedback will result in higher error rates since users have less indications of the pen's contact with the board.
H.5. Writing on the slanted board will be easier from a physical workload standpoint.

5.2 Participants

We have recruited 43 participants in local universities and 3 participants are excluded from the study: 2 due to physical discomfort (eye strain and back pain) and 1 due to distraction (answering his phone multiple times during the experiment). Out of the N = 40 participants considered, there are 26 males and 14 females aged between 20-67 (M = 29.1, SD = 10.4). There are 9 participants who use VR regularly and 31 who used VR once at most. A minimum of Common European Framework of Reference for Languages (CEFRL) B1 English level is required to participate. No participants are native English speakers. A total of 38 participants are right-handed and 2 are left-handed.

5.3 Apparatus and materials

The user study run on a laptop with an Intel Core i7 CPU and dedicated NVIDIA GeForce RTX 2070 Super GPU. The program is developed in Unity 2020.3.17.f1 using C# scripting. The WritePad SDK [47] is used for the handwriting recognition and is integrated into the Unity3D application. We use a HTC Vive Pro 2 HMD and Steam Index controllers.

5.4 Task and procedure

The experiment uses a 2×2 within-subject design. The independent variables are the orientation of the writing board (2 levels: vertical and slanted) and the use of sensory feedback (2 levels: all feedback and no feedback). Condition order is controlled with a balanced latin square. The task is to write 10 phrases that are randomly generated in advance from the MacKenzie phrase set. Phrases remain visible during the writing task since no participants are native english speakers and this can help reduce error rates [25]. A particular attention is put into generating phrases sets that are representative of the MacKenzie set by generating sets that fall in a standard deviation range of measures such as words per phrase, etc. 32 phrases are excluded from the set because they either contain proper nouns or words that cannot be found in the English dictionary such as "jedi"; or are deemed to difficult to understand for non-native English speakers. Phrases can be found in Supplemental Materials. Phrases cannot appear in multiple sessions and all participants write the same phrases in the same order. Each condition has a total of 53-55 words (238-246 characters) to write. Only the order of the conditions changes per participant. All participants perform 4 writing sessions (one per condition, which order is counterbalanced). Before each session and writing task, participants are instructed to write as naturally, as fast and as accurately as possible.

Participants first fill out a consent form and a demographics questionnaire before performing a baseline handwriting task on paper. It consists of writing 10 predetermined phrases randomly chosen from the MacKenzie phrase set. This task is timed from the first contact of the pen on paper till the end of the writing session. A training session in VR is then carried out where participants write one phrase on each board (first the vertical one then the slanted one) and get familiar with the controls. The training phrase contains all alphabet characters so participants can find out if they need to adapt their handwriting according to the recognizer's abilities. During training, participants are also asked to configure the slanted board's height and angle, see Fig. 3 so that they write comfortably without having to raise their arm, to avoid physical strain. The effect of these parameters on performance is not studied as it depends on user preferences and morphology, and is not enforced nor counterbalanced for statistical analysis. This is discussed in Sect. 6. The goal is mainly to maintain a similar posture between participants to keep physical demand consistent. For example, we want to prevent mismatches between one's height and the board's height. Most participants have chosen a high (N = 33) and as vertical as possible (N = 25) slanted board. More details can be found in Fig. 3. A second training is then carried out for participants to understand how the phrases are displayed, see Fig. 1, how they can choose between recognition alternatives if needed and how they can correct the previous word if the recognition did not work properly. During the whole experiment, correction is not forced i.e., users can choose to correct or not an erroneous word, as the literature has shown that there is no significant differences in typing speed and total error rate between forced and recommended correction [1]. Additionally, we wanted to prevent participants from getting stuck on a word if there was some incompatibility between one's handwriting and the recognizer. This second training consists of writing 3 phrases. In total, during both training phases, participants

have to write 5 phrases. The main study is then carried out in a virtual classroom. Participants are asked to write 10 phrases. After each writing session, participants fill out the System Usability Scale (SUS) and the NASA Task Load Index (NASA-TLX) questionnaires. There are 4 writing sessions in total per participant as there are two different writing boards and two different sensory feedback types. Participants write 40 phrases in total. At the end, participants are able to freely comment the experiment and give feedback. The experimenter also explains the goal of the experiment which lasts about 1 hour in total. A video of the experiment can be found in Supplemental Materials.

With 40 participants, each writing 10 phrases per condition, a total of 4 conditions, we collect $40 \times 10 \times 4 = 1600$ phrases. Training phrases are not analyzed. This experiment is approved by Nantes University's ethics committee with number 17112022.

5.5 Measures

This section states the different measures used in this study and defines some new metrics adapted to word-level systems.

5.5.1 WPM

For measuring text entry speed, we use the WPM metric (1). However, in our case, the timer starts as soon as the pen gets in contact with the board, which means the first character written is taken into account. As such, the definition we use for measuring entry speed is the following:

$$WPM = (|T|/S) \times 60 \times (1/5)\,. \tag{5}$$

Additionally, the timer runs from the first contact between the pen and the board after a phrase is displayed until the validation of the phrase's last word. The timer stops after each phrase and resumes after the next contact to let participants read phrases before writing them [1]. Since the system uses a word-by-word entry paradigm, we automatically add spaces after each validated word in the transcribed string. Since validating a word requires a button press, it is equivalent to pressing the space bar on a keyboard.

5.5.2 Error rates

In this study, we use MSD ER (4) and TER (3) to report error rates. MSD ER is used over ER as it is a robust and well defined way to compute error rate. Error rate metrics work at a character-level and are used for comparison purposes with other text entry systems. However, since our system works at a word-level, we need to introduce new metrics that better represent the performance of the word recognition system. Adapted Word Error Rate (WER) and Total Word Error Rate (TWER) are defined as follows:

$$WER = (INF_w/|T_w|) \times 100\,, \tag{6}$$

$$TWER = (INF_w + IF_w)/(C_w + INF_w + IF_w) \times 100\,, \tag{7}$$

where INF_w is the number of Incorrect Not Fixed words, $|T_w|$ is the number of words in the transcribed string, IF_w is the number of Incorrect Fixed words, C_w is the number of Correct words.

Finally, since the system lets users choose between three recognition alternatives after each word is validated, we design a metric for keeping track of the ratio between alternatives words and total written words, called Alternative Rate (AR):

$$AR = A_w/(C_w + INF_w + IF_w) \times 100\,, \tag{8}$$

where A_w is the number of times an alternative is chosen. Indeed, choosing an alternative is only considered an error if the alternative itself is wrong. Being able to choose between alternatives is part of the design rationale of the system so correct choices do not count as errors nor corrections.

5.5.3 Usability and workload

Usability is measured with the SUS questionnaire [6]. Results can be compared to other systems and interpreted with descriptive adjectives [3]. Higher scores means higher usability and vice versa. Workload is measured with the NASA-TLX questionnaire that aims to quantify physical, mental and temporal demand, subjective performance, frustration and effort [18]. Lower scores are better for NASA-TLX.

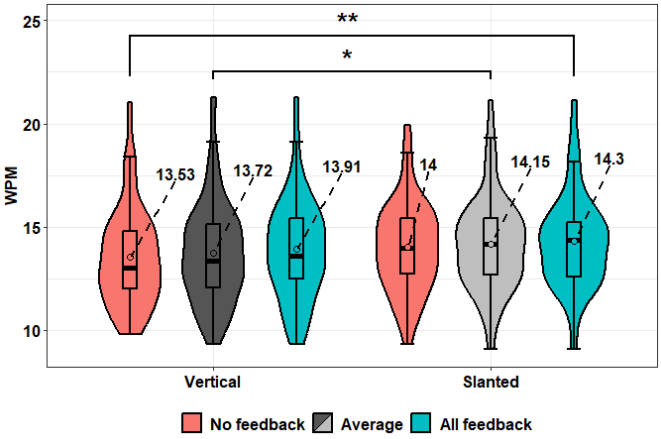

Fig. 4: Writing speeds (WPM) across board and feedback types. Dark (resp. light) grey shows results averaged over the vertical (resp. slanted) board. Blue (resp. Red) shows results with (resp. without) sensory feedback. Circles show average.

5.6 Results

We use a two-way repeated measures ANOVA and Tukey's HSD correction for pairwise comparisons. Since there are only two levels per dependant variable, no correction for violations of the sphericity assumption is used. The two independent variables are 'board', which represents the orientation of the board (either vertical or slanted) and 'feedback'. This part describes the results of this analysis.

5.6.1 Text entry speed

Fig. 4 shows writing speeds across all conditions and averaged for each board. ANOVA results show that 'board' ($F_{1,39} = 7.164$, $p = 0.011$, $\eta_p^2 = 0.155$) has a significant effect on text entry speed. We find that the slanted board has higher WPM than the vertical one. Participants reached on average 14.15 WPM (SD = 2.26) on the slanted board while they reached a significantly lower 13.72 WPM (SD = 2.45) on average on the vertical board. No significant effect of 'feedback' ($F_{1,39} = 3.944$, $p = 0.054$, $\eta_p^2 = 0.092$) is found on writing speed. No significant interaction effect is found. The overall average text entry speed is 13.94 WPM (SD = 2.36). The best overall performance is reached on the slanted board with sensory feedback to attain 14.3 WPM (SD = 2.32). In the baseline condition, participants reach an average 30.36 WPM (SD = 5.02) when writing with a pen on paper.

5.6.2 Error rate

Fig. 5 shows average TWER across all conditions, ER and WER averaged for each feedback type. ANOVA results show no significant main or interaction effect on TER. The overall average TER is 9.85% (SD = 7.1). When using the slanted board and sensory feedback, which is the best condition in regards to entry speed, average TER is 9.28% (SD = 6.92). There is a significant main effect of 'feedback' ($F_{1,39} = 6.01$, $p = 0.019$, $\eta_p^2 = 0.134$) on ER. With sensory feedback, participants have an average 1.05% (SD = 1.86) ER which is significantly lower than the 1.63% (SD = 2.96) reached without any sensory feedback. There is no significant main effect of 'board' nor any interaction effect on ER. The overall average ER is 1.34% (SD = 2.48). Similar results are found on WER, with a significant main effect of 'feedback' ($F_{1,39} = 7.665$, $p = 0.009$, $\eta_p^2 = 0.164$). With sensory feedback, participants have an average 1.7% (SD = 2.65) WER which is significantly lower than the 2.54% (SD = 3.90) reached without any sensory feedback. The overall average WER is 2.12% (SD = 3.35). There is a significant main effect of 'feedback' ($F_{1,39} = 8.198$, $p = 0.007$, $\eta_p^2 = 0.174$) on TWER. We find that with sensory feedback TWER is significantly lower (M = 13.85%, SD = 8.49) than without any feedback (M = 15.65%, SD = 9.43). No significant main effect of 'board' nor

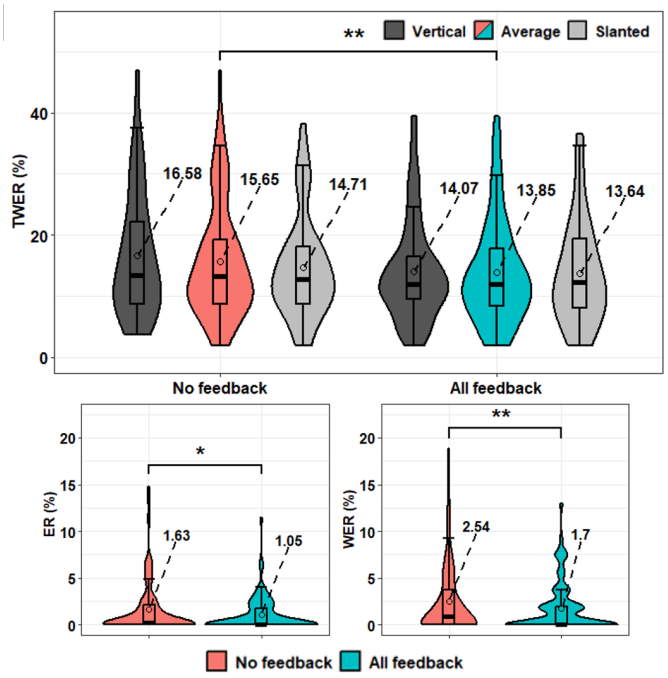

Fig. 5: TWER across all conditions averaged on feedback type (top) and influence of feedback on ER and WER (bottom). Dark (resp. light) grey shows results for the vertical (resp. slanted) board. Blue (resp. red) shows results averaged with (resp. without) sensory feedback.

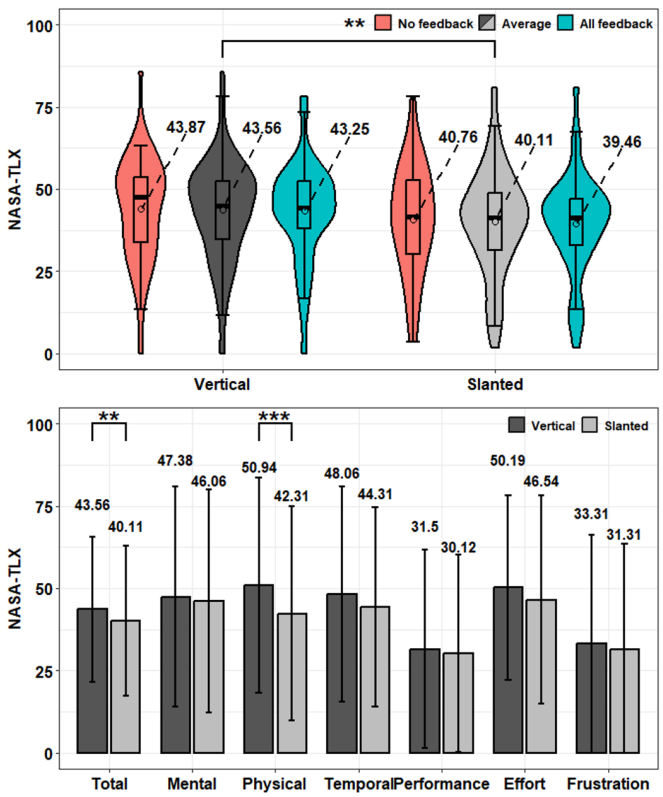

Fig. 6: NASA-TLX results across board and feedback type (top) and NASA-TLX profiles for the board types (bottom). Dark (resp. light) grey shows results averaged over the vertical (resp. slanted) board. Blue (resp. red) shows results with (resp. without) sensory feedback.

interaction effect are found. The overall average TWER is 14.75% (SD = 8.99). There is one extreme outlier in TWER in the vertical board with sensory feedback condition (TWER = 39.5%) that is not removed during analysis because no evidence supports technical issues. There are no significant main or interaction effects on AR. The overall average AR is 10.39% (SD = 5.74).

5.6.3 Usability

The ANOVA shows no significant main effect nor interaction effect on SUS results. The overall average SUS result is 80.05 (SD = 12.67) which places the system in the "Good" to "Excellent" range [3]. In the vertical board condition with feedback (resp. without), SUS score is 79.75 (resp. 78.38). In the slanted board condition with feedback (resp. without), SUS score is 80.88 (resp. 81.19).

5.6.4 Workload

Fig. 6 shows NASA-TLX results across all conditions and averaged for each board type. It also shows complete NASA-TLX results when using the slanted board. ANOVA results show a significant main effect of 'board' ($F_{1,39} = 9.759$, $p = 0.003$, $\eta_p^2 = 0.200$) on NASA-TLX results, but no significant main effect of 'feedback' nor any interaction effect. Participants rate the workload significantly lower when using the slanted board (M = 40.11, SD = 16.09) over the vertical board (M = 43.56, SD = 15.62). Overall NASA-TLX average score is 41.84 (SD = 15.90). There is a significant main effect of 'board' ($F_{1,39} = 17.821$, $p < 0.001$, $\eta_p^2 = 0.314$) on NASA-TLX physical workload component, but no significant main effect of 'feedback' nor any interaction effect. Participants rate the physical workload significantly lower with the slanted board (M = 42.31, SD = 22.97) over the vertical board (M = 50.94, SD = 23.13). No other significant effect is found on other components of NASA-TLX.

5.6.5 Learnability

We conducted a one-way repeated measures ANOVA with Tukey's HSD correction for pairwise comparisons to study learnability. Sphericity is verified by a Mauchly's test ($p = 0.167$). Fig. 7 shows the evolution over 4 sessions of (left) entry speed for all participants based on their experience with VR and (right) the evolution of ER for all participants. ANOVA results show a significant main effect of 'session' ($F_{3,117} = 8.197$, $p < 0.001$, $\eta_p^2 = 0.879$) on WPM when considering all participants. Pairwise comparisons show significant differences between sessions $1vs3$, $1vs4$, $2vs3$, $2vs4$. Overall average writing speed after 10 phrases is 13.48 WPM and raises to 14.5 WPM after 40 phrases. For VR experienced participants, average writing speed after 10 phrases is 14.93 WPM and raises to 16.16 WPM after 40 phrases. A similar study is conducted on all error rates but no effects are found.

5.6.6 VR familiarity

We conduct a two-way mixed repeated measures ANOVA with Tukey's HSD correction for pairwise comparisons to study the impact of VR experience on entry speed. The between independent variable is 'experience' and has 2 levels : high and low extracted from answers to the demographic questionnaire. Low experience means having used VR at most one time. The within independent variables are 'feedback' and 'board'. ANOVA results show there is a significant main effect of 'experience' ($F_{1,38} = 6.919$, $p = 0.012$, $\eta_p^2 = 0.154$), a significant main effect of 'feedback' ($F_{1,38} = 5.223$, $p = 0.028$, $\eta_p^2 = 0.121$) but no significant main effect of 'board' on writing speed. Users with low VR experience reach on average 13.49 WPM (SD = 1.88) while users with VR experience reach a significantly higher 15.46 WPM (SD = 3.12).

6 Discussion

6.1 Handwriting results

Average entry speed across all participants for handwriting after 40 training phrases is 14.5 WPM. Users with VR experience reach 16.16 WPM after the same training. Using a slanted board is best for writing speed as it comes with a significantly reduced physical demand as shown by NASA-TLX physical score of 42.31 against 50.94 on the

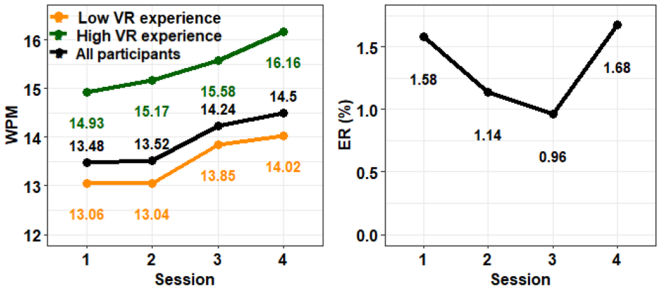

Fig. 7: The effect of learning on WPM (left) and on ER (right). Black lines show results for all participants, green (resp. orange) lines show results for participants with high (resp. low) VR experience.

vertical board. We observe a trend on the effect of sensory feedback on entry speed that might be confirmed with more participants or a finer study of feedback. The best overall writing condition is the slanted board with sensory feedback which lets users reach an average of 14.3 WPM and 9.28% TER. Overall TER is 9.85% and there is no learning effect on any error metric. The board type does not have any effect on any error metric. However, the presence of sensory feedback reduces ER, WER and TWER. We hypothesize the presence of sensory feedback gives users more indications of the contact between pen and board which reduces unwanted strokes. Sensory feedback also helps reduce TWER, which is not visible on TER. This can be explained by how the recognizer works: it is more prone to errors on short words since it is easier to mistake them for other words as less data is available for recognition. Moreover, average English words length is low which means there are more candidates for shorter words. Errors are mainly observed in short words, with a low amount of characters and errors on short words have greater impact on TWER than on TER. Character and word-level metrics are thus not to be used interchangeably since they do not show the same results but word-level metrics provide additional insight into the performance of word-based systems. This is especially true for techniques where performance varies depending on the length of words. On average, participants rely on recognition alternatives 10.39% of the time. The system is deemed usable with a "Good" rating of 80.05 on the SUS questionnaire. NASA-TLX results show the system has high subjective performance (M = 30.81) and low frustration (M = 32.31), with a moderately high physical demand (M = 42.31) that is comparable to virtual keyboard techniques.

Results support our hypotheses **H.4.** and **H.5.**. Sensory feedback indeed helps reduce ER, WER, and TWER, moreover physical workload is lower when using the slanted writing board. However, **H.1.**, **H.2.** and **H.3.** are not supported by our results. While users reported that writing on the vertical board is indeed more familiar, the higher physical workload of this condition makes it less effective when it comes to writing speed. There is no effect of using a slanted board on any error rate metric. Finally, the effect of sensory feedback on writing speed is not shown during this study, however a non-significant trend ($p = 0.054$, $\eta_p^2 = 0.092$) suggests a more refined study of feedback might show effects on writing speed. Some participants report being overwhelmed by the sounds and vibrations when sensory feedback is activated. Disabling only one of those would maybe yield better results. The experimental design does not allow for analysis of the influence of slanted board height and angle which depends on user preference. Participants are instructed to find comfortable writing conditions without having to raise their arm, to maintain a coherent physical demand across participants. Fig. 3 shows the disparity in group sizes especially when it comes to board height. Condition order is not balanced in each group. No statistical trends of the effect of height and angle appear on any metric but further work would be required to draw any conclusions.

The design rationale of our system is based around the fact that adult handwriting speed is said to be around 40 WPM [42]. In our baseline observations, participants reach an average of 30.36 WPM for a 10 MacKenzie phrases task when writing on paper. While the evaluation of handwriting speed is different, this means the performance ceiling of the system is potentially lower than expected. However, it is important to remember that participants are not native English speakers which might have also impacted handwriting speed on the baseline condition and in VR. A similar effect is observed in PizzaText with differences in typing speed for native and non-native Latin alphabet users [52].

As stated in the system's description, we introduce a visuo-proprioceptive conflict when simulating collisions between the pen and the writing board. We believe this conflict had little to no impact on performance as no participant report having noticed the conflict. Additionally, when collision is activated — as part of the sensory feedback condition — we notice participants usually write at the limit of losing contact between pen and board. This means in most cases the displacement between the real and virtual hand is small.

The study shows the efficiency of handwriting as a text entry method in VR. We believe however that our system still has room for improvement. In its current iteration, editing written text can be challenging as one must rewrite whole words. The editing process must be improved to provide an optimal user experience. Moreover, the recognizer is trained on English handwriting but tested on non-native English writers. There are differences in cursive letters between different Latin alphabet languages which makes certain words hard to recognize. From our observations, a consequent proportion of the registered errors are due to participants getting stuck on particular words, even if they had the possibility to leave them uncorrected. For participants that reported a high frustration on the NASA-TLX questionnaire, this was often the main culprit. In real-life applications, we would recommend using a recognizer that can learn the handwriting of its users, which is something we did not study here.

6.2 Analysis of results from the literature

According to Sect. 2, we compare our system with state-of-the-art controller-based virtual text entry techniques that are compatible with our industrial use case such as the raycast keyboard [4], the drum-like keyboard [4], HiPad [20] and Flower Text Entry [27]. We decide not to compare our entry speed and error rate measures to Speicher et al. [40] results for the raycast keyboard (called controller pointing) because they do not use the same phrase set and they only write 5 phrases per condition. However, since Boletsis et al. [4] did not study NASA-TLX scores, we compare them against Speicher et al. [40] raycast keyboard and controller tapping. Indeed, there is no available evaluation of NASA-TLX scores on the drum-like keyboard to the best of our knowledge and controller tapping seems like a good-enough approximation for indirect comparison. We also discuss the recent study of handwriting in VR by Venkatakrishnan et al. [44].

There are many differences in the demographics of each study which should be taken into consideration. Our 40 participants, only 9 of which have notable VR experience, are on average older (M = 29.1, SD = 10.4) than in other studies. With our target of industrial VR applications, we believe our population is more representative of this use case. In [4], a raycast vs. drum-like keyboards study has 22 participants (8 with major VR experience), aged on average 25.77. HiPad has 15 participants aged on average 22.5 and no precision is given regarding VR experience. Flower Text Entry has 10 participants aged between 21 and 25. All of them have VR experience and 5 of them have already participated in previous pilot studies. By having participated in previous studies, these participants have closer to 240 phrases of training than the 80 phrases claimed in the discussion. As such, it seems fair to compare our VR experienced users to novice users of Flower Text Entry. For the comparison with the raycast vs. drum-like keyboards study, we have a higher proportion of inexperienced VR users.

Regarding entry speed, handwriting reaches 14.02 WPM for inexperienced VR users and 16.16 WPM for experienced VR users after 40 phrases of training. Overall average entry speed after 40 phrases is 14.5 WPM across all participants. Raycast keyboard users reach an average of 16.65 WPM after 10 phrases. Drum-like keyboard users reach an average of 21.01 WPM after 10 phrases. However, these results come from Boletsis et al. [4] which used only one phrase set to evaluate 4 techniques, meaning those results might be slightly over-

evaluated as participants were able to learn the phrases. HiPad users reach 13.57 WPM after 60 phrases. Flower Text Entry novice users, to be compared with our VR experienced users, reach 13.05 WPM after $\sim$40 training phrases (after Day 3) and attain 17.65 WPM after $\sim$80 phrases. In writing conditions similar to ours, participants in Venkatakrishnan et al.'s [44] study reach $\sim$10.5 WPM on a random selection of the 200 most commonly used English words. The higher writing speed displayed in our study can be explained by multiple reasons: writing random words instead of coherent phrases might be more difficult and we use a similar principle than the best writing condition studied in [44] i.e., direct contact of a virtual pen on a planar virtual board, but we refine board orientation and sensory feedback. Moreover, our validation and erase interactions that use a single controller button press rather than a VE interaction are likely more efficient as they do not require any arm movement. Compared to HiPad [20], our method is also faster after less training. While Flower Text Entry's authors [27] conclude that their method is faster than raycast and drum-like keyboards, we believe that training time is a major factor that must be taken into account. The performance of these keyboards is evaluated after 10 phrases which makes them faster, after equivalent training, than our handwriting method (13.48 WPM) or Flower Text Entry (8.96 WPM). However, raycast and drum-like keyboards are dual-handed techniques unlike the other compared techniques. To the best of our knowledge, there is no available evaluation of one-handed versions of these keyboards which makes comparison difficult [4, 10, 40]. Keyboards have the advantage of having great performance from the get-go but it would be interesting to compare their long-term performance to the other techniques, as Flower Text Entry's novice but VR experienced users surpass raycast keyboards (with only 10 training phrases) after 80 training phrases. Since keyboards are a very familiar way to input text, their performance might be capped earlier than techniques that rely on different principles and there are no long-term performance studies to the best of our knowledge [4, 40]. We believe the speed of handwriting in VR is still raising after 40 training phrases even if results show no significant difference between sessions 3 and 4. We suspect that participants were less precise when writing words, thus requiring more corrections, in the fourth session which they knew was the last one considering our experiment was quite long (1h on average). This is supported by the fact that ER is higher in the fourth session than in the third one, and is comparable to the first session.

Regarding error rates, handwriting's ER is 1.05% with sensory feedback and TER is 9.85% overall across all participants. For the raycast and drum-like keyboards, TER is respectively 11.05% and 12.11% [4]. HiPad's ER is 0.22% and TER is 4.94%. Flower Text Entry has low ER (0.09%) and TER (2.50%). Both HiPad and Flower Text Entry have lower total error rates than handwriting. Our Handwriting TER is comparable with that of raycast and drum-like keyboards. However, handwriting TER is in most cases not due to user error, as they write the correct words, but to recognition errors which means using a better handwriting recognition tool could greatly improve the TER, and by extension entry speed of handwriting in VR. On the other hand, errors in raycast and drum-like keyboards, HiPad, or Flower Text Entry errors are mainly due to users. Handwriting recognition can already rival the accuracy of state-of-the-art text entry methods, and has the potential to become even more accurate in the future. Indeed, with a $\sim$12% corrected word error rate (difference between TWER and WER) and a $\sim$10% AR (the rate at which users relied on recognition alternatives), our system has a $\sim$22% failure rate on main recognition proposition. In the Venkatakrishnan et al.'s [44] study on VR handwriting, the recognizer used [21] is claimed by its developers to have only 10.4% WER and 4.3% ER. This performance is actually similar to that claimed by WritePad SDK [47], which is the recognizer we used. However, Venkatakrishnan et al. [44] do not clearly state how handwriting accuracy is computed – it seems the authors only look at ER – and it is hard to say whether the claimed recognizer performance is actually reached.

As for workload, the overall NASA-TLX score (the lower the better) for handwriting on a slanted board is 40.11 with a 42.31 physical workload score, a 30.12 subjective performance score, and a 31.31 frustration score. In Speicher et al. [40], the controller pointing keyboard, similar to the raycast keyboard in [4], has a 37.86 physical workload score, a 28.10 frustration score, and a 28.33 subjective performance score. The controller tapping keyboard [40], which we believe can be used as an approximation of the drum-like keyboard for physical workload, has a 51.90 physical workload score. Flower Text Entry reports a low overall workload of 28.06 with audio and tactile feedback in a QWERTY layout in a pilot user study. No detail is given about the different NASA-TLX items such as physical workload. Our handwriting entry exhibits similar workloads as the controller pointing keyboard, even in the physical workload category. It performs better than a controller tapping keyboard in terms of physical demand which may be due to the fact that our system is one-handed. According to NASA-TLX results, handwriting is a viable alternative to virtual keyboards for text entry in VR regarding performance and frustration, which were the main items of the design rationale, without compromise on physical demand.

6.3 Limitations

This study has known limitations. First, it would be interesting for the baseline comparison to be done on an actual whiteboard, maybe even wearing an HMD with pass-through vision. This would give an indication of the real performance ceiling of the handwriting task in VR. Indeed, after writing 40 phrases, participants are still improving, even if the frequently changing writing conditions hinders learnability of the system. Thus, the performance ceiling is not quantified. Moreover, the Mackenzie task is vastly different from industrial data entry scenarios that we designed the system for. For example, symbols, numbers, and special characters are not studied. A direct comparison of handwriting to state-of-the-art data entry methods in VR on a more industry-focused task is needed. Additionally, this study is focused on standing up handwriting with no passive haptic feedback which corresponds to our industrial use case. Other scenarios, such as using tangible surfaces for passive haptic feedback could be studied for other practical applications of VR. Finally, in this study, sensory feedback is made of visual, haptic and auditory feedback that are all turned off or on simultaneously. A more refined study of sensory feedback could confirm or disprove the observed trend of sensory feedback on writing speeds.

7 Conclusion

In this paper, we present a handwriting system to input textual data in VR with a focus on industrial applications. Our technique uses handwriting recognition to transform handwritten words on a virtual board into exploitable textual data. Using a single controller, we rely on the playfulness of writing in VR to achieve high typing speed and high adoptability. The system is designed around 3 main criteria: efficiency, ease-of-learning and familiarity. This makes it a suitable tool for inputting textual data in industrial settings that require efficiency and low frustration for repetitive tasks, as proven by a usability and workload study. Our user study shows that after 40 phrases, users reach an average writing speed of 14.5 WPM with 9.3% error rate. The highest observed speed is 21.1 WPM in a 10-phrase writing task. The physical demand is comparable with the most commonly used virtual keyboards techniques such as head or controller pointing [40], while frustration is low and subjective performance is high. Handwriting is a great alternative to virtual keyboards that could be further improved by using more efficient handwriting recognition to lower error rate thus increasing writing speeds. Future work will reevaluate the system after recognition improvements by comparing our system with state-of-the-art data entry methods in an industrial task that better represents our envisioned use-case. An evaluation of our system and virtual keyboards learnability and performance ceilings should be performed over multiple sessions.

Supplemental Materials

All supplemental materials are available on OSF at `https://doi.org/10.17605/OSF.IO/NBV2G`, released under a CC BY 4.0 license. In particular, they include (1) .csv files of the used and removed MacKenzie phrases and (2) a video that describes the different experimental conditions.

REFERENCES

[1] A. S. Arif and W. Stuerzlinger. Analysis of text entry performance metrics. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, pp. 100–105, Sept. 2009. doi: 10.1109/TIC-STH.2009.5444533

[2] K. Aryal, A. Badamasi, U. U.M., and D. Mansir. Drivers and Barriers of Virtual Reality Adoption in UK AEC Industry. *Engineering Construction & Architectural Management*, Mar. 2021. doi: 10.1108/ECAM-09-2020-0685

[3] A. Bangor. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*, 4(3), 2009.

[4] C. Boletsis and S. Kongsvik. Controller-based Text-input Techniques for Virtual Reality: An Empirical Comparison. *International Journal of Virtual Reality*, 19(3):2–15, Oct. 2019. Number: 3. doi: 10.20870/IJVR.2019.19.3.2917

[5] D. A. Bowman, C. J. Rhoton, and M. S. Pinho. Text Input Techniques for Immersive Virtual Environments: An Empirical Comparison. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(26):2154–2158, Sept. 2002. doi: 10.1177/154193120204602611

[6] J. Brooke. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, Nov. 1995.

[7] E. Chang, H. T. Kim, and B. Yoo. Virtual Reality Sickness: A Review of Causes and Measurements. *International Journal of Human–Computer Interaction*, 36(17):1658–1682, Oct. 2020. doi: 10.1080/10447318.2020.1778351

[8] S. Chen, J. Wang, S. Guerra, N. Mittal, and S. Prakkamakul. Exploring Word-gesture Text Entry Techniques in Virtual Reality. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, pp. 1–6. Association for Computing Machinery, New York, NY, USA, May 2019. doi: 10.1145/3290607.3312762

[9] S. Dash. BlueTap — The Ultimate Virtual-Reality (VR) Keyboard, Apr. 2017.

[10] T. J. Dube and A. S. Arif. Text Entry in Virtual Reality: A Comprehensive Review of the Literature. In M. Kurosu, ed., *Human-Computer Interaction. Recognition and Interaction Technologies*, Lecture Notes in Computer Science, pp. 419–437. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-22643-5_33

[11] J. Dudley, H. Benko, D. Wigdor, and P. O. Kristensson. Performance Envelopes of Virtual Keyboard Text Input Strategies in Virtual Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 289–300, Oct. 2019. ISSN: 1554-7868. doi: 10.1109/ISMAR.2019.00027

[12] R. Elmgren. Handwriting in VR as a Text Input Method. Master Thesis, KTH, 2017.

[13] A. Erickson, K. Kim, G. Bruder, and G. Welch. Effects of Dark Mode Graphics on Visual Acuity and Fatigue with Virtual Reality Head-Mounted Displays. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, May 2020. doi: 10.1109/VR46266.2020.1580695145399

[14] Z. Fu, J. Xu, Z. Zhu, A. X. Liu, and X. Sun. Writing in the Air with WiFi Signals for Virtual Reality Devices. *IEEE Transactions on Mobile Computing*, 18(2):473–484, Feb. 2019. doi: 10.1109/TMC.2018.2831709

[15] G. González, J. P. Molina, A. S. García, D. Martínez, and P. González. Evaluation of Text Input Techniques in Immersive Virtual Environments. In J. A. Macías, A. Granollers Saltiveri, and P. M. Latorre, eds., *New Trends on Human-Computer Interaction*, pp. 1–10. Springer London, London, 2009. doi: 10.1007/978-1-84882-352-5_11

[16] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Text Entry in Immersive Head-Mounted Display-Based Virtual Reality Using Standard Keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 159–166, Mar. 2018. doi: 10.1109/VR.2018.8446059

[17] Y.-S. Han, J. Lee, J. Lee, W. Lee, and K. Lee. 3D CAD data extraction and conversion for application of augmented/virtual reality to the construction of ships and offshore structures. *International Journal of Computer Integrated Manufacturing*, 32(7):658–668, July 2019. doi: 10.1080/0951192X.2019.1599440

[18] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, vol. 52, pp. 139–183. Elsevier, 1988. doi: 10.1016/S0166-4115(08)62386-9

[19] C.-H. Hsu, C.-H. Chung, R. Venkatakrishnan, R. Venkatakrlshnan, Y.-S. Wang, and S. V. Babu. Comparative Evaluation of Digital Writing and Art in Real and Immersive Virtual Environments. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 1–10. IEEE, Lisboa, Portugal, Mar. 2021. doi: 10.1109/VR50410.2021.00089

[20] H. Jiang and D. Weng. HiPad: Text entry for Head-Mounted Displays Using Circular Touchpad. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 692–703, Mar. 2020. ISSN: 2642-5254. doi: 10.1109/VR46266.2020.00092

[21] D. Keysers, T. Deselaers, H. A. Rowley, L.-L. Wang, and V. Carbune. Multi-Language Online Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1180–1194, June 2017. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. doi: 10.1109/TPAMI.2016.2572693

[22] M. j. Kim, X. Wang, H. Li, and S.-C. Kang. Virtual Reality for the Built Environment: A Critical Review of Recent Advances. *Journal of Information Technology in Construction*, 18:279–305, Aug. 2013.

[23] S. Kim and D. Weissmann. Middleware-based Integration of Multiple CAD and PDM Systems into Virtual Reality Environment. *Computer-Aided Design and Applications*, 3(5):547–556, Jan. 2006. doi: 10.1080/16864360.2006.10738408

[24] P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–9. Association for Computing Machinery, New York, NY, USA, Apr. 2018.

[25] P. O. Kristensson and K. Vertanen. Performance comparisons of phrase sets and presentation styles for text entry evaluations. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pp. 29–32. ACM, Lisbon Portugal, Feb. 2012. doi: 10.1145/2166966.2166972

[26] T. Law, D. Lam, J. Endang, and M. Chan. How rad(-ical) is VRAD (Virtual Reality-Aided Design)? In *2020 24th International Conference Information Visualisation (IV)*, pp. 751–756, Sept. 2020. ISSN: 2375-0138. doi: 10.1109/IV51561.2020.00133

[27] J. Leng, L. Wang, X. Liu, X. Shi, and M. Wang. Efficient Flower Text Entry in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, Aug. 2022. Publisher: IEEE Computer Society. doi: 10.1109/TVCG.2022.3203101

[28] X. Lu, D. Yu, H.-N. Liang, W. Xu, Y. Chen, X. Li, and K. Hasan. Exploration of Hands-free Text Entry Techniques For Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 344–349, Nov. 2020. ISSN: 1554-7868. doi: 10.1109/ISMAR50242.2020.00061

[29] I. S. MacKenzie and R. W. Soukoreff. Phrase Sets for Evaluating Text Entry Techniques. *NEW HORIZONS*, p. 2, 2003.

[30] I. S. MacKenzie and K. Tanaka-Ishii. *Text entry systems: mobility, accessibility, universality*. The Morgan Kaufmann series in interactive technologies. Boston Morgan Kaufmann, Amsterdam, 2007.

[31] D. Morais, M. Waldie, and D. Larkins. The Evolution of Virtual Reality in Shipbuilding. *Proceedings of COMPIT 2017*, p. 12, 2017.

[32] N. Nooruddin, R. Dembani, and N. Maitlo. HGR: Hand-Gesture-Recognition Based Text Input Method for AR/VR Wearable Devices. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 744–751, Oct. 2020. ISSN: 2577-1655. doi: 10.1109/SMC42975.2020.9283348

[33] R. Outerelo. Meta-review of text input approaches within VR - A study on the platform's viability as a productivity workspace. Master Thesis, Universitat Oberta de Catalunya, 2020.

[34] S. Pick, A. S. Puika, and T. W. Kuhlen. SWIFTER: Design and evaluation of a speech-based text input metaphor for immersive virtual environments. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 109–112, Mar. 2016. doi: 10.1109/3DUI.2016.7460039

[35] I. Poupyrev, N. Tomokazu, and S. Weghorst. Virtual Notepad: handwriting in immersive VR. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)*, pp. 126–132, Mar. 1998. doi: 10.1109/VRAIS.1998.658467

[36] V. Rajanna and J. P. Hansen. Gaze typing in virtual reality: impact of keyboard design, selection method, and motion. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, pp. 1–10. ACM, Warsaw Poland, June 2018. doi: 10.1145/3204493.3204541

[37] J. Shen, J. Hu, J. J. Dudley, and P. O. Kristensson. Personalization of a Mid-Air Gesture Keyboard using Multi-Objective Bayesian Optimization. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 702–710. IEEE, Singapore, Singapore, Oct. 2022. doi: 10.1109/ISMAR55827.2022.00088

[38] Z. Song, J. J. Dudley, and P. O. Kristensson. Efficient Special Character Entry on a Virtual Keyboard by Hand Gesture-Based Mode Switching. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 864–871. IEEE, Singapore, Singapore, Oct. 2022. doi: 10.1109/ISMAR55827.2022.00105

[39] R. W. Soukoreff and I. S. MacKenzie. Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. In *Conference on Human Factors in Computing Systems*, p. 8, 2003.

[40] M. Speicher, A. M. Feit, P. Ziegler, and A. Krüger. Selection-based Text Entry in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. Association for Computing Machinery, New York, NY, USA, Apr. 2018.

[41] R. Stark, J. H. Israel, and T. Wöhler. Towards hybrid modelling environments—Merging desktop-CAD and virtual reality-technologies. *CIRP Annals*, 59(1):179–182, Jan. 2010. doi: 10.1016/j.cirp.2010.03.102

[42] J. Summers and F. Catarro. Assessment of handwriting speed and factors influencing written output of university students in examinations. *Australian Occupational Therapy Journal*, 50(3):148–157, 2003. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1440-1630.2003.00310.x. doi: 10.1046/j.1440-1630.2003.00310.x

[43] Y. Tang and H. Gu. CAD Model's Simplification and Conversion for Virtual Reality. In *2010 Third International Conference on Information and Computing*, vol. 4, pp. 265–268, June 2010. ISSN: 2160-7451. doi: 10.1109/ICIC.2010.338

[44] R. Venkatakrishnan, R. Venkatakrishnan, C.-H. Chung, S. V. Babu, and Y.-S. Wang. Investigating a Combination of Input Modalities, Canvas Geometries, and Inking Triggers on On-Air Handwriting in Virtual Reality. *ACM Transactions on Applied Perception*, p. 3560817, Sept. 2022. doi: 10.1145/3560817

[45] K. Vertanen. *Efficient Correction Interfaces for Speech Recognition*. PhD thesis, University of Cambridge, Cambridge, 2009.

[46] J. Whyte, N. Bouchlaghem, A. Thorpe, and R. Mccaffer. *From CAD to virtual reality: Modeling approaches, data exchange, and interactive 3-D building design tools.*, vol. 10 of *Automation in Construction*. Elsevier, 2000.

[47] WritePadSDK. https://github.com/phatware/WritePadSDK.

[48] W. Xu, H.-N. Liang, Y. Zhao, T. Zhang, D. Yu, and D. Monteiro. RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1991–2001, May 2019. doi: 10.1109/TVCG.2019.2898736

[49] N. Yanagihara and B. Shizuki. Cubic Keyboard for Virtual Reality. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*, SUI '18, p. 170. Association for Computing Machinery, New York, NY, USA, Oct. 2018. doi: 10.1145/3267782.3274687

[50] X. Yi, C. Yu, M. Zhang, S. Gao, K. Sun, and Y. Shi. ATK: Enabling Ten-Finger Freehand Typing in Air Based on 3D Hand Tracking Data. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 539–548. ACM, Charlotte NC USA, Nov. 2015. doi: 10.1145/2807442.2807504

[51] C. Yu, Y. Gu, Z. Yang, X. Yi, H. Luo, and Y. Shi. Tap, Dwell or Gesture? Exploring Head-Based Text Entry Techniques for HMDs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 4479–4488. Association for Computing Machinery, New York, NY, USA, May 2017. doi: 10.1145/3025453.3025964

[52] D. Yu, K. Fan, H. Zhang, D. Monteiro, W. Xu, and H.-N. Liang. PizzaText: Text Entry for Virtual Reality Systems Using Dual Thumbsticks. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2927–2935, Nov. 2018. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2018.2868581

[53] Z. Zhang, M. Sun, B. Gao, and L. Wang. 2-Thumbs Typing: A Novel Bimanual Text Entry Method in Virtual Reality Environments. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 530–531, Mar. 2021. doi: 10.1109/VRW52623.2021.00147